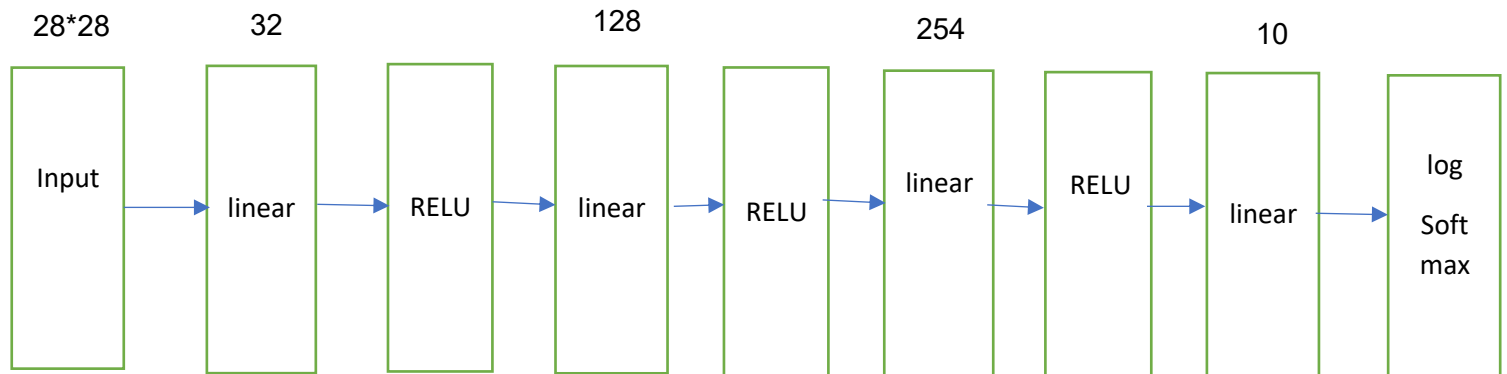


תרגיל בית 1

הרשת בנויה סה"כ משמונה שכבות (7 שכבות חבויות). ארכיטקטורת הרשת מורכבת משכבות ליניאריות ו-RELU לסירוגין כאשר השכבה האחרונה הינה log softMAX. וכל שכבה מחוברת באופן מלא לשכבה שקודמת לה. (סה"כ פרמטרים: 64660)

ניתן לתאר את הרשת באופן וויזואלי כך:

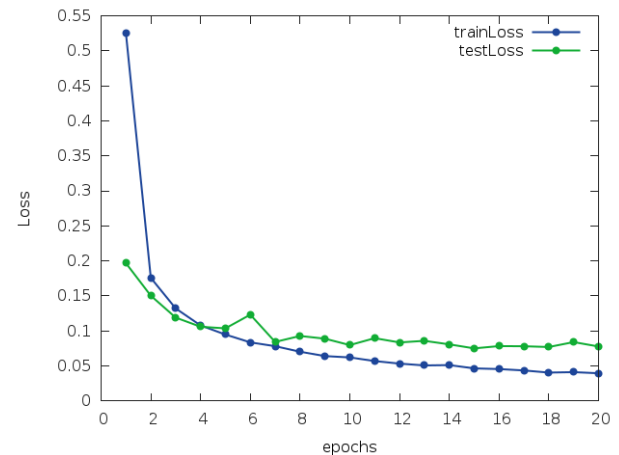
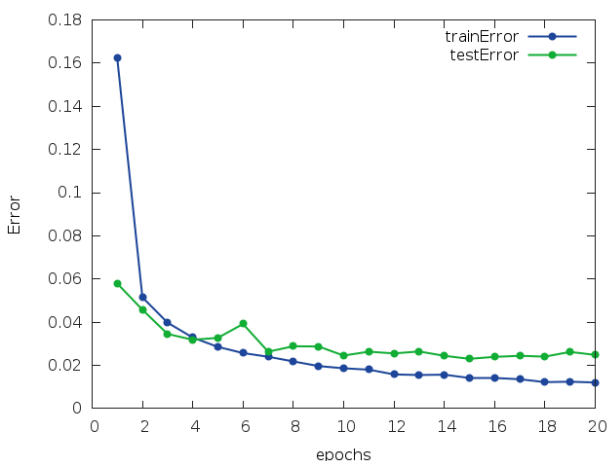


הנתונים מנורמלים לפני תהליך האימון, גם סט-האימון וגם סט-המבחן מנורמלים לפי סט-האימון. כלומר, כל אחד מהפיקסלים מנורמל ע"י חיסור ממוצע הפיקסלים בסט-הנתונים וחלוקה בסטיית התקן. לפני כל epoch ביצענו shuffle של סט האימון.

```
Training error: 0.012086004273504    Training Loss: 0.039678894449025
Test error: 0.024839743589744    Test Loss: 0.07784884251081
```

התוצאות שהתקבלו:

### פונקציית האופטימיזציה שהשתמשנו בה היא Stochastic Gradient Descent

נסיונות:

- שימוש בפונקציית אופטימיזציה adagrad (ללא שינוי מספר ה-epochs) אך ללא שיפור בשגיאת המבחן (0.0291) ובשגיאת האימון (0.0214).
- שימוש בפונקציית אופטימיזציה SGD ושינוי מספר ה-epochs ל-30. לא חל שיפור בשגיאת המבחן (0.0252) וחל שיפור בשגיאת האימון (0.0096).

- שימוש בפונקציית אופטימיזציה SGD ושינוי מספר epochs ל-5. לא חל שיפור בשגיאת המבחן (0.0354) וגם לא בשגיאת האימון (0.0309).
- שימוש בפונקציית אופטימיזציה SGD ושינוי מספר batch'ים ל-16. לא חל שיפור בשגיאת המבחן (0.0147) ובשגיאת האימון (0.0327).
- שימוש בפונקציית אופטימיזציה SGD ושינוי מקדם הלמידה מ-0.1 ל-0.2. שגיאת האימון נשארה כפי שהייתה (0.0129) ולא חל שיפור בשגיאת המבחן (0.0293).

#### מסקנות:

1. על פי ניסיונותינו, פונקציית adagrad נותנת תוצאות פחות טובות מפונקציות הsgd עבור מסד הנתונים עליו עבדנו.
  2. כאשר מגדילים את מספר epochs ישנו חשש לover fitting – כפי שהתקבל בניסיונותינו, קיבלנו כי שגיאת האימון קטנה בצורה משמעותית, ואילו שגיאת המבחן עלתה במעט מהשגיאה המקורית.
  3. כפי שצפינו, כאשר מקטינים את מספר epochs שגיאת המבחן וכן שגיאת האימון גדלות (מכיוון שמבצעים פחות איטרציות של האימון- המערכת מתקשה ללמוד בצורה טובה את המידע).
  4. על פי ניסיונותינו, לא מצאנו קשר ישיר בין שינוי מספר batch'ים/ מקדם הלמידה לבין שיפור בשגיאות המבחן והאימון, על כן בחרנו אותם באמצעות tuning.
- לבסוף כתבנו את הפונקציה predict שמקבלת את שם המודל כפרמטר ומחזירה את השגיאה. נציין, כי פונקציה זו מנרמלת את הנתונים לפי ערכי סט-האימון (כפי שהתבצע בשלב אימון הרשת)